A Packet Loss Concealment Method Using Recursive Linear Prediction

Kazuhiro Kondo and Kiyoshi Nakagawa

Department of Electrical Engineering,
Yamagata University, Japan
{kkondo,nakagawa}@yz.yamagata-u.ac.jp

Abstract

We proposed and evaluated a speech packet loss concealment method, which predicts lost segment from speech included in both packets before and after the lost packet. The lost segments are predicted by recursively using linear prediction both in the forward direction from the packet preceding the loss, as well as in the backward direction from the packet succeeding the lost segment. The predicted sample in each direction is smoothed to obtain the final interpolated signal. The adjacent segments are also smoothed to significantly reduce the discontinuity between the interpolated signals. Subjective quality of the proposed method showed higher scores than the packet loss concealment algorithm described in the ITU standard G.711 Appendix I, with MOS rating exceeding 2.4, even at an extremely high packet loss rate of 30%.

1. Introduction

In many of the real time speech communication systems that transmit signals using packets, *e.g.* the emerging VOIP systems, there exists a finite probability that some packets will be lost. The two major sources for packet loss are congestion in the intermediate nodes, and discarded packets at the receiving end due to packets arriving too late to be decoded and played out. These packet losses may in some cases lead to intervals with significant loss rate, requiring a robust concealment solution.

Various methods have been proposed to try to generate an estimate of the lost speech segments included in the lost packets. Some require modifications to the data included in the packets, *i.e.* integration of overhead information to help the accurate estimation of the lost segments, while others attempt this only from the speech signal in the previously received packets. We will only deal with the latter in this paper.

The simplest estimate of the lost segment is to substitute the lost signal with silence or gain-adjusted pseudo-random noise. However, this has been known to degrade the perceived speech quality significantly. The next best solution is to simply copy the last packet, and substitute a replica of the speech signal included in the copied packet. This provides somewhat less degradation, but in many cases is still not sufficient.

Recently, ITU has standardized a packet loss concealment (PLC) algorithm to be used with the G.711 PCM standard [1]. This algorithm uses pitch detection to estimate the best matching pitch period immediately after the last received data, and repeats the pitch period data to fill the lost segment. This algorithm provides reasonably good estimate of the lost segment at fairly low complexity. An enhanced algorithm, which performs pitch period repetition in the LPC residual domain, has been proposed and standardized by ANSI [2],[3]. The proposed enhancement showed modest improvement over the method described in [1].

In this paper, we propose a PLC algorithm that predicts lost speech segment from speech included in both packets before and after the lost packet. Recursive linear prediction is employed both in the forward direction, *i.e.* from the packet preceding the loss, as well as in the backward direction, *i.e.* from the packet succeeding the lost segment. The predicted sample in each direction is smoothed to obtain the final interpolated signal. Smoothing is also applied to the adjacent segments to significantly reduce the discontinuity between the interpolated signals. We compared the subjective quality of our algorithm to the G.711 Appendix I PLC algorithm. The results show that the two algorithms show similar quality up to a packet loss rate of about 10%, but the proposed algorithm generally shows better quality above this packet loss rate.

In the next section, we describe the proposed algorithm. In section 3, the subjective quality evaluation test as well as its results is shown. Finally, the conclusion is given in section 4.

2. Packet loss concealment algorithm using recursive LPC

2.1. Concealment of a single lost packet

Figure 1 shows the configuration of the proposed algorithm. All processing is done at the receiving end only. We will use linear prediction recursively to try to estimate the lost

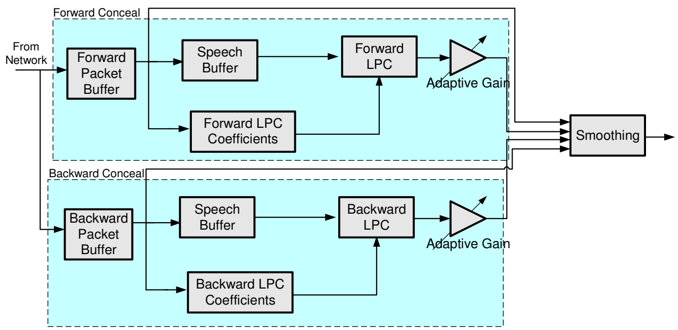

Figure 1. Block diagram of the proposed PLC algorithm

speech segment from speech samples in the neighboring received packets. We can write the forward prediction of one sample from the preceding received packet using the standard linear prediction equation [4]:

$$\hat{x}_{f,n} = -1 * \sum_{i=1}^{N} a_{i+1} x_{n-i}$$

where $\hat{x}_{f,n}$ is the predicted sample, x_{n-i} is the previous speech sample, N is the LPC prediction order, and a_i is the LPC coefficient calculated from $x_{n-i}, i = 1,..M$ where M is the analysis window length. The following samples, $\hat{x}_{f,n+i}$, are predicted recursively using predicted samples with previous speech samples. For example, both

$$\hat{x}_{f,n}$$

and

$$x_i\ ,\ i = n-1,\ n-2,\ \ldots\ ,\ n-N+1$$

are used to predict sample $\hat{x}_{f,n+1}$. In this case, the LPC coefficients a_i are not updated; they remain fixed at values predicted from $x_{n-i}, i = 1,..M$. This sample prediction is repeated for the whole lost segment, *i.e.* $\hat{x}_{f,n+i}, i = 0,1,..L-1$, where L is the lost segment length.

As the prediction is repeated, the gain of the predicted speech was shown to gradually decrease. Thus, we introduced a linearly increasing gain G_f starting from 1.0 at the start of the lost segment, and saturating at G_{max} at the end of the segment. This gain is applied to the predicted speech samples. We used an empirical value of 1.8 for G_{max}.

The above prediction can be applied in backwards as well using the packet received after the lost segment. Following the above notation, this can be written as:

$$\hat{x}_{b,n} = -1 * \sum_{i=1}^{N} b_{i+1} x_{n-i+N+1}$$

where b_i is the backward LPC coefficient, and $\hat{x}_{b,n}$ is the backward predicted sample. This prediction is repeated to obtain the whole lost segment in time-reversed order. Adaptive gain G_b is also applied in a similar manner, this time starting from 1.0 at the end of the lost segment, to G_{max} at the start of the segment in the backward direction. Obviously, in order to predict $\hat{x}_{b,n+i}$, we need to receive the packet after the lost packet, which adds to the processing delay.

We now have two estimate for the lost segment, *i.e.* $\hat{x}_{f,n+i}$ and $\hat{x}_{b,n+i}$. It can reasonably be assumed that the former ($\hat{x}_{f,n+i}$) is a better estimate of the earlier portions of the lost segment since the recursive repetition of the LPC estimation is smaller, while the latter ($\hat{x}_{b,n+i}$) is a better estimate of the later portions. Thus, we can combine the two estimates with a linear weight to obtain a single sample estimate:

$$\hat{x}_{n+i} = (1-\alpha) \cdot \hat{x}_{f,n+i} + \alpha \cdot \hat{x}_{b,n+i}$$

where α is a linearly increasing weight from 0 for i = 0, to 1 for $i = L - 1$.

2.2. Consecutive lost packets

When consecutive packets are lost, the processing will depend on the combination of normal reception or loss of both the preceding and succeeding packet.

(1) **Preceding packet received, succeeding packet lost:** In this case, only forward prediction will be used. LPC coefficients are calculated from the samples in the preceding packet. Lost segment is recursively calculated using the above LPC coefficients and samples in the preceding packet.

(2) **Preceding packet lost, succeeding packet lost:** In this case, only forward prediction will be used. LPC coefficients are kept fixed at values calculated from speech samples in the last received packet.

(3) **Preceding packet lost, succeeding packet received:** In this case, only backward prediction will be used. Backward LPC coefficients are calculated from samples in the succeeding packet. Lost segment is calculated using the backward LPC coefficients and the samples in the succeeding packet recursively in the backward direction.

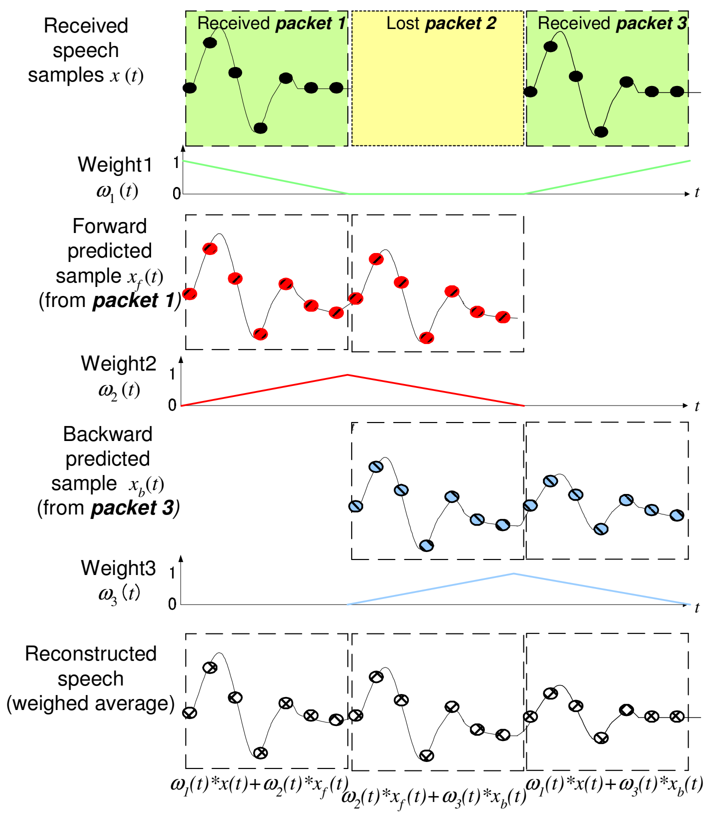

Figure 2. Smoothing between predicted and received samples

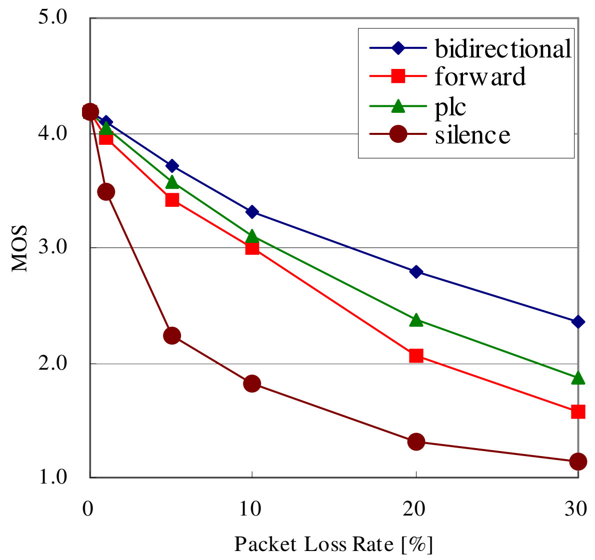

Figure 3. Mean opinion score vs. packet loss rate for clean speech

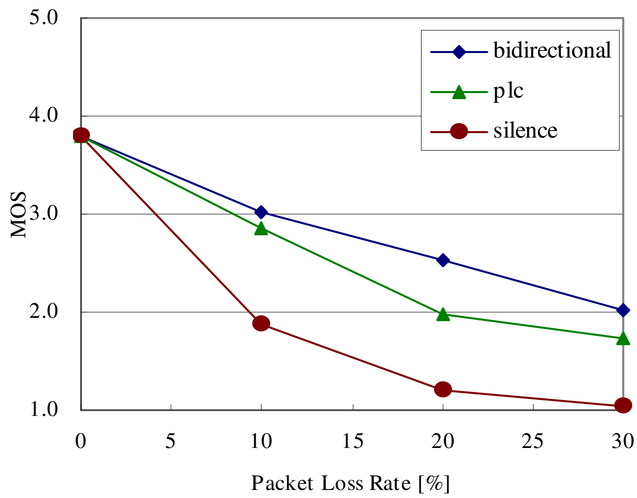

Figure 4. Mean opinion score vs. packet loss rate for G.729 transcoded speech

2.3. Smoothing of adjacent received packets

In most cases, the predicted lost segments show some degree of discontinuity with the adjacent received samples. This seems to be a major source of perceived quality degradation. Thus, transition smoothing from received speech to the predicted samples is essential to reduce this degradation. An example of smoothing for the single lost packet case is shown in Fig. 2.

In order to smooth transition from received samples to the predicted samples, a forward-predicted sample is prepared for the received packet preceding the lost packet (***packet 1***). The LPC coefficients used here are calculated from the samples in the same packet, *i.e.* ***packet 1***. However, the samples included in the packet before ***packet 1*** are used to predict these samples. This predicted segment is again linearly weighed and added to the actual received segment in ***packet 1***; larger weight is given to the predicted segment at samples in the smoothed segment just before the lost segment, while larger weight is given to the received samples at the beginning of the smoothed segment.

The same smoothing can be applied to the samples after the lost segment, *i.e.* ***packet 3***. Backward-predicted samples for the segment in ***packet 3*** are weighed and averaged with received samples. This obviously requires more processing delay since two packets after the loss needs to be received in this case. Thus, in the listening tests in section 3, we decided not to use the backward-prediction smoothing described here. However, use of this smoothing does seem to improve the perceived quality to some degree.

Smoothing is also applied to consecutively lost segments. For example, the overlapped forward and backward predicted samples are averaged to smooth the boundary between the forward-predicted samples and backward-predicted samples. Extensive smoothing described here has shown to improve the subjective quality significantly in informal listening tests.

3. Subjective quality comparison tests

We conducted subjective quality evaluation tests for the proposed algorithm. Twenty-one listeners with normal hearing participated in the test. We used samples in the ASJ continuous speech corpus [5],[6]. Two male and two female speakers, two samples each, were used. The sample consisted of read phonetically balanced Japanese sentences. Each was approximately 3 seconds long. The original sampling rate was 16k Hz, but was down-sampled to 8 kHz. All samples were in 16 bit linear PCM.

Packet length was assumed to be 10 [msec], or 80 samples at 8 kHz. Packets were randomly discarded. The tested packet loss concealment schemes were as follows:

(1) Simple silence insertion (denoted **silence** in the results).

(2) ITU-T G.711 Appendix I (denoted **plc**). The pitch search range was set between 50 and 1 kHz, which is broader than is written in the standard. We also did not use the 2-phase coarse-fine search, but used a single-phase fine search.

(3) The proposed algorithm. We tested both the bidirectional LPC (denoted **bidirectional**) based prediction as well as prediction in the forward direction only (denoted **forward**). An LPC order of 128, with analysis window length of 256 samples was used. One-sided Hamming window was applied in the LPC analysis. As stated in 2.3, we did not use the backward prediction smoothing, but all other smoothing modes described in 2.3 were applied.

Figure 3 shows the mean opinion score for packet loss rate between 0 and 30 % with clean speech. As expected, silence insertion shows significant degradation with loss. The other methods show similar trends up to about 10 %

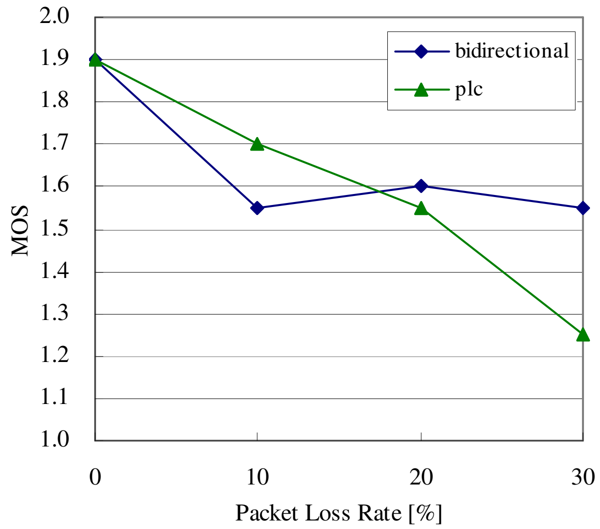

Figure 5. Mean opinion score vs. packet loss rate for speech with babble

loss. However, above this rate, bidirectional LPC outperforms the G.711 PLC and the forward LPC. The G.711 PLC start to show synthetic quality due to long pitch repetition for high loss rates. On the other hand, the forward LPC shows some "background ringing" perhaps due to over-enhancement of the formant due to the long recursive LPC. The introduction of bandwidth expansion [7] may reduce this effect.

We also conducted subjective quality tests with degraded speech inputs. Figure 4 shows the MOS vs. packet loss rate for speech coded and decoded using the ITU-T G.729 codec. This kind of setup may occur in VOIP or wireless networks including cellular networks, where speech may go through a speech coding and decoding in order to reduce the transmission bit rate requirement. This test was intended to find out the effect of synthetic artifacts introduced by the codec on the concealment algorithm. The results seem similar to Fig. 3, but the difference between bidirectional LPC and G.711 PLC is smaller. This may be due to the fact that the G.729 codec shows some low-pass characteristics, which consequently serves to make the perceived degradation due to pitch repetition less apparent.

Figure 5 shows MOS vs. packet loss rate for speech mixed with babble (multi-speaker noise). This test was intended to find out the effect of surrounding speech noise on the concealment algorithm. Babble was added to the clean speech sample at 10 dB SNR. We only tested for one female speech sample, and compared the concealed speech quality for bidirectional LPC with the PLC algorithm. Again, the bidirectional LPC seems to maintain better quality than the PLC at higher loss rates. The repetition of pitch period with babble seems to show synthetic quality with PLC, which seems to degrade the perceived quality.

4. Conclusions and future work

In this paper, we proposed a speech packet loss concealment algorithm, which uses LPC recursively to estimate the lost segment. When both forward LPC from the packet preceding the lost packet and backward LPC from the packet succeeding is used, the subjective quality was shown to outperform the packet loss concealment in ITU-T G.711 Appendix I [1]. The proposed algorithm showed superior quality especially at loss rates above 10%.

The algorithm requires a high LPC order to be effective. Thus, the computational complexity is fairly high. However, use of block update or gradient update may still be effective while considerably lowering the complexity of the LPC coefficient update.

The bidirectional LPC shows improved loss concealment performance, but will require additional processing delay since it requires the speech signal in the packet succeeding the lost packet. However, it was shown that for loss rates up to 10%, the subjective quality is not significantly different from forward only LPC, which does not require the added delay. Thus, it may be possible to limit the overall delay by using a bimodal prediction scheme based on the observed packet loss rate; when the loss rate is below 10%, the forward LPC is used, while for loss rates above 10%, the LPC is switched to bidirectional. This should provide a balanced solution between the speech quality and processing delay. The switch between modes can be accomplished during inter-word pauses fairly transparently. Expanding or shrinking the pause slightly can also accomplish this delay adjustment. There have been reports that the human ear is immune to such alterations if the percentage of the alterations is limited to a small proportion [8].

5. References

[1] ITU-T Recommendation G.711 Appendix I, "A High Quality Low Complexity Algorithm for Packet Loss Concealment with G.711," ITU-T, 1999.

[2] ANSI Recommendation T1.521a-2000 (Annex B), "Packet Loss Concealment for Use with ITU-T Recommendation G.711," ANSI, 2000.

[3] E. Gündüzhan and K. Momtahan, "A Linear Prediction Based Packet Loss Concealment Algorithm for PCM Coded Speech," *IEEE Transaction on Speech and Audio Processing*, vol. 9, no. 8, pp. 778-785, Nov. 2001.

[4] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Upper Saddle River, NJ, 1996.

[5] Japan Information Processing Development Corporation, *ASJ Continuous Speech Corpus for Research*, vol. 1-3, Aug. 1991.

[6] S. Itahashi, "On Recent Speech Corpora Activities in Japan," *Journal of the Acoustical Society of Japan (E)*, vol. 20, no. 3, pp. 163-170, 1999.

[7] Y. Tohkura et. al., "Spectral Smoothing Technique in PARCOR Speech Analysis-synthesis," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-26, pp. 587-596, Dec. 1978.

[8] W. Montgomery, "Techniques for Packet Voice Synchronization," *IEEE Journal on Selected Areas in Communications,* vol. SAC-1, no. 6, pp. 1022-1028, Dec. 1983.